Overview of OpenSRS

TOWARD AN OPEN IMPLEMENTATION OF ISO 11238

Tyler Peryea, Noel Southall, Trung Nguyen

February 4, 2013

NCATS

NIH National Center for Advancing Translational Sciences

The proposed implementation is a cultivation of our efforts over the years to *develop*, *demonstrate*, and *disseminate* software for managing, mining, and visualizing chemical biology data.

Outline

- ISO 11238 and scope
- Reference substance database
- Software requirements
- Implementation roadmap
- Architecture overview
- Technology stack
- Status & milestones
- Discussion (and demo)

ISO 11238 and Scope

- Substance categories
  - Chemical
  - Protein
  - Nucleic acid
  - Polymer
  - Structurally diverse
- Specified substances Groups 1–4
- Official names in multiple languages, jurisdictions, and domains
- Well-defined references and relationships between substances
- Unique identifiers

Reference Substance Database

- A reference substance database is distributed with each OpenSRS deployment
  - » Bootstrap from FDA's public SRS records
- A public accessible "master" copy of the database is housed at NCATS
  - » Data curation
  - » Conflict resolution
- Defined update schedule

Software Requirements

- Self-contained and modular
  - » Run entirely on a desktop or access remotely
- Well-defined data access application programming interface (API)
  - » Mobile or third-party clients
- Fine-gained security model
  - » Access control for every piece of information
- Audit trail of all data fields
- Web- and desktop-based clients
  - » Multiple platforms (e.g., Linux, Windows, Mac) for desktop client
- Basic support for text, structure, sequence searching
- Wizard interface to guide registration
- Support attachments (e.g., PDF's, MS spectra, images)
- Configurable "business rules" for standardizing structures

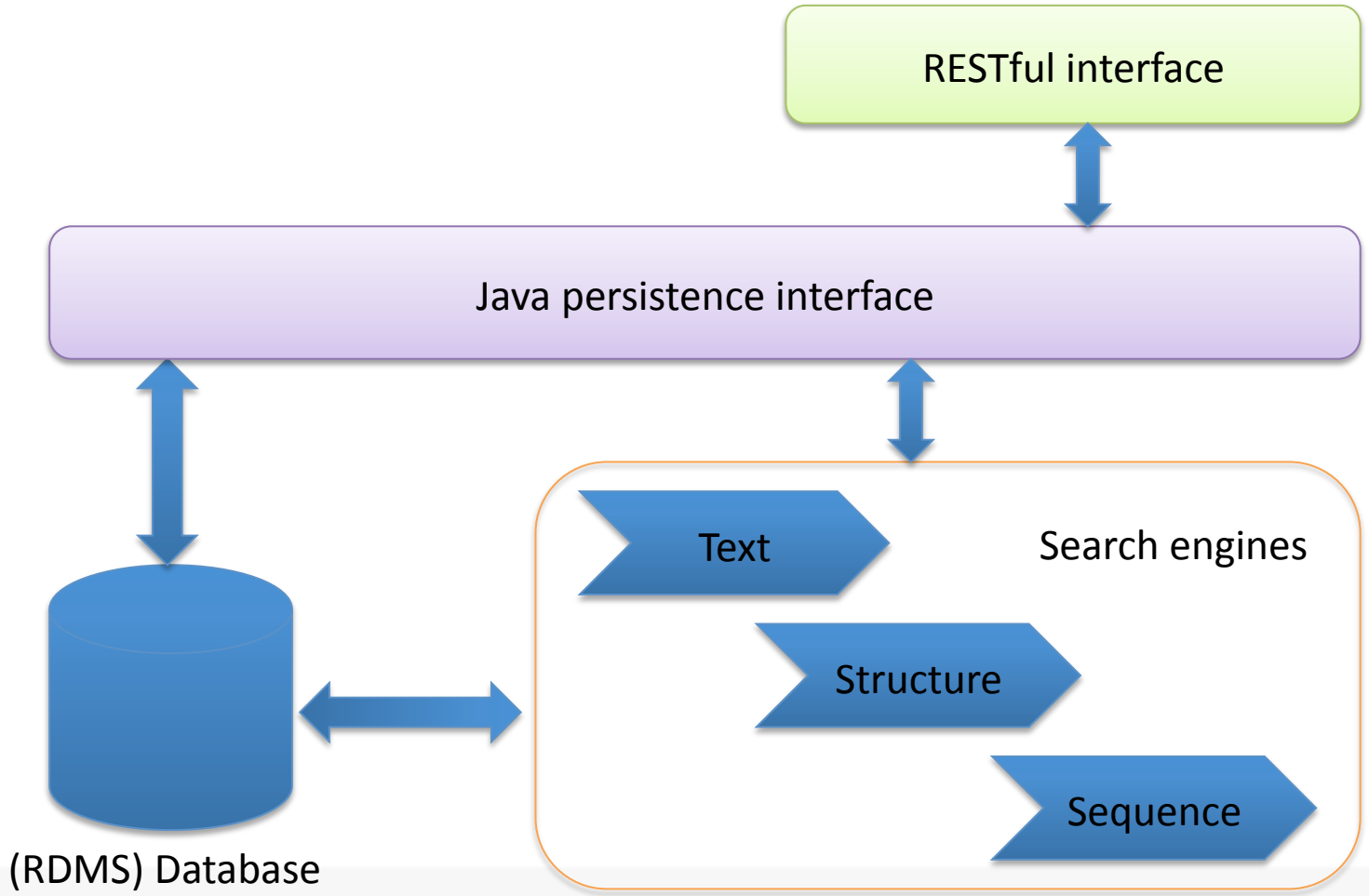
Implementation Roadmap

- Multiple stages
- Stage I (alpha)
  - » Use cases limited to FDA & CBG-MEB
  - » Solidify backend and data models
  - » Chemical substance
- Stage II (beta)
  - » Limited deployment at FDA
  - » Support polymer & protein substances
- Stage III (beta)
  - » Support for remaining substance categories
  - » ISO 11238 compliance
  - » Open deployments to other organizations
- Stage IV (public release)

Architecture Overview

- Client-server
  - » Server is self-hosting when installed on desktop
- Backend database agnostic (e.g., Oracle, MySQL)
- Support two APIs: Native (Java) and RESTful
- Standalone server or deployable within a standard web container (e.g., Glassfish, Tomcat)
- Pluggable chemical toolkit
  - » Default toolkit is a redistributable version of JChem
- Pluggable engines for text, structure, and sequence searching

Architecture Overview (cont'd)



Technology Stack

- Web-based client
  - » Combination of client- and server-side technologies (e.g., ExtJS, JSF)
- Desktop client
  - » Java Swing and other open source libraries
  - » Deploy as either signed webstart or installed image
- Server
  - » JDO as the persistence layer
  - » Lucene text search engine
  - » Custom implementations of structure and sequence search engines
  - » Standalone server based on embedded Jetty or Glassfish and H2 database

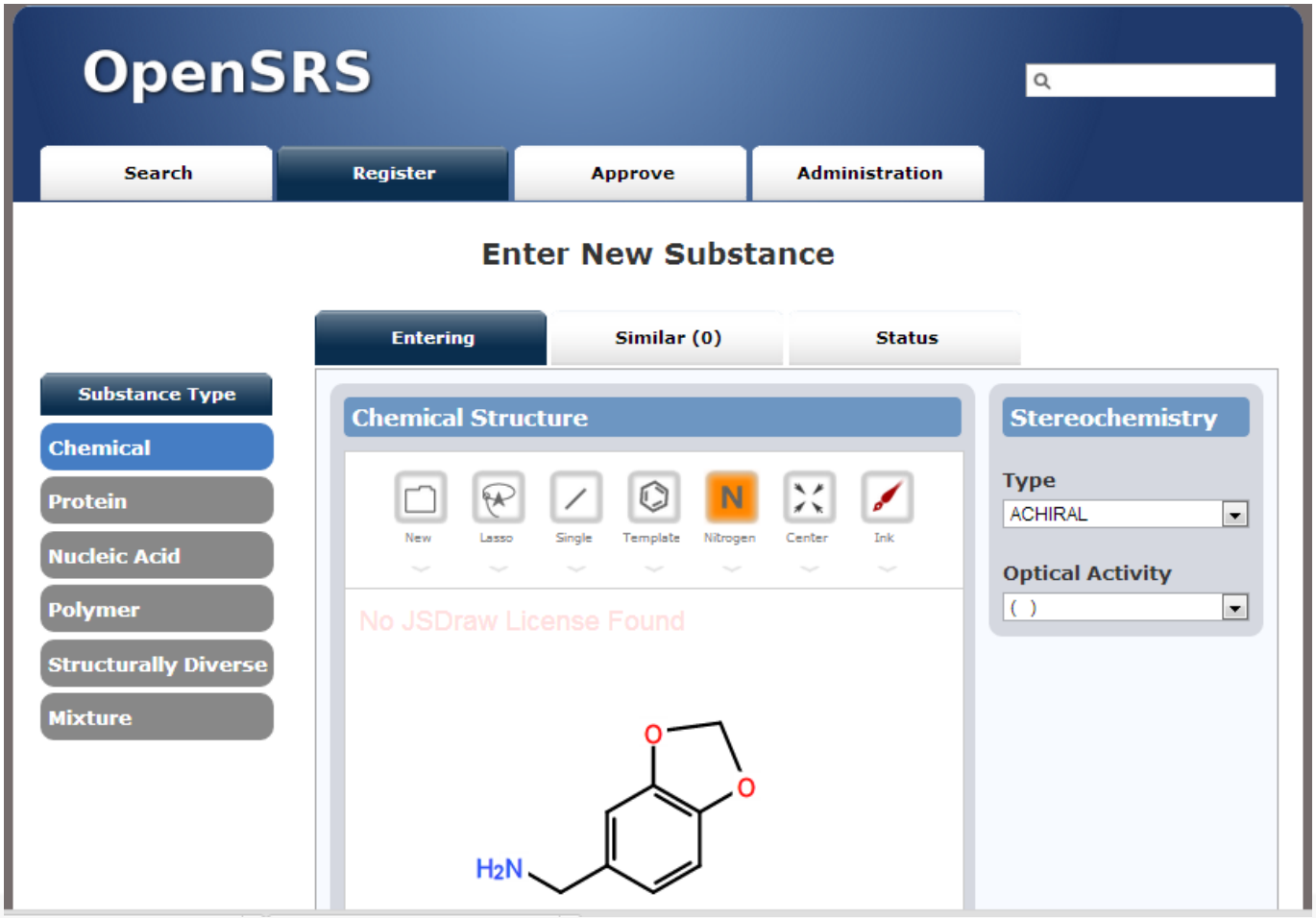
Status & Milestones

- Basic user interface framework for web and desktop clients in place
- On-going development on the server backend
  - » Solidify the Java persistence API
  - » Develop initial set of “business rules”
  - » RESTful API
- Anticipate completion of Stage I in June 2013
- If all goes well, Stage II by early 2014

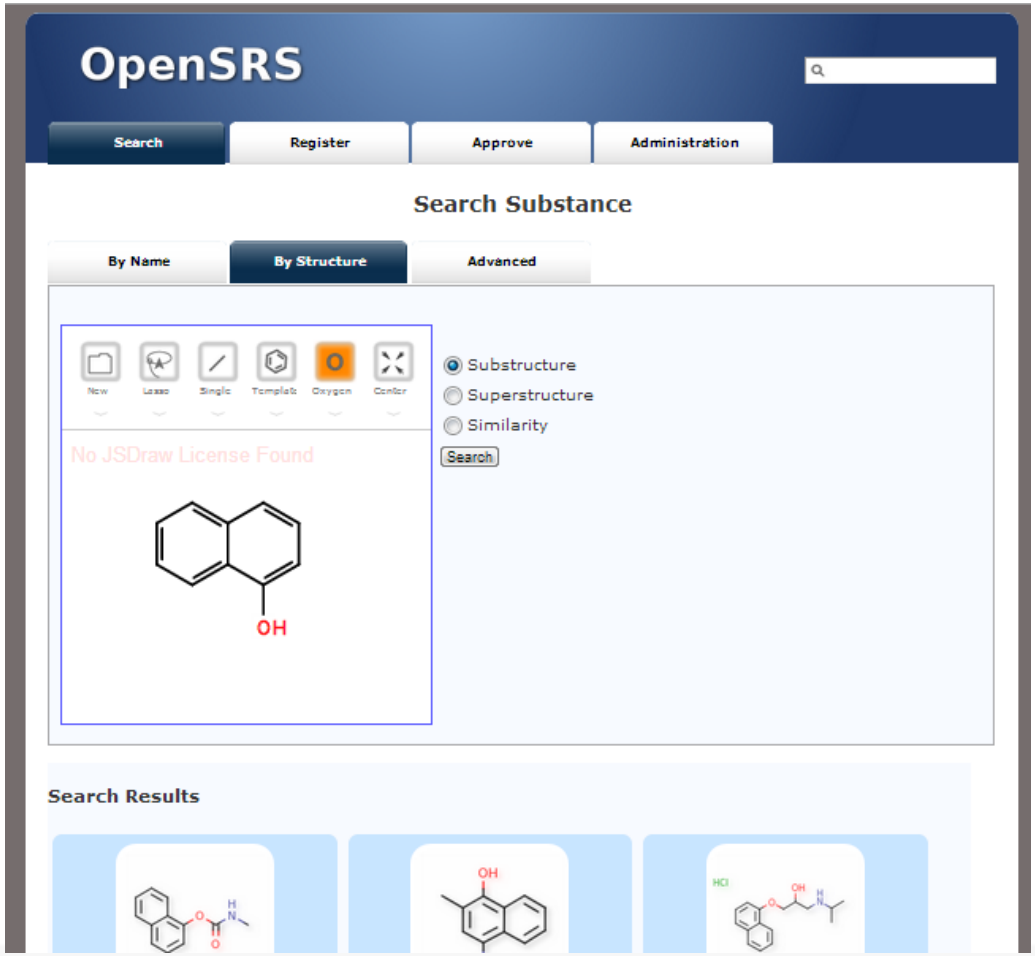
Discussion

- Organic development model
  - » Features evolve through iterations
- Code is hosted at bitbucket.org
  - » Repository is currently private but access is available upon request
  - » Repository goes public in Stage II (beta)
- Demos available for web and desktop clients

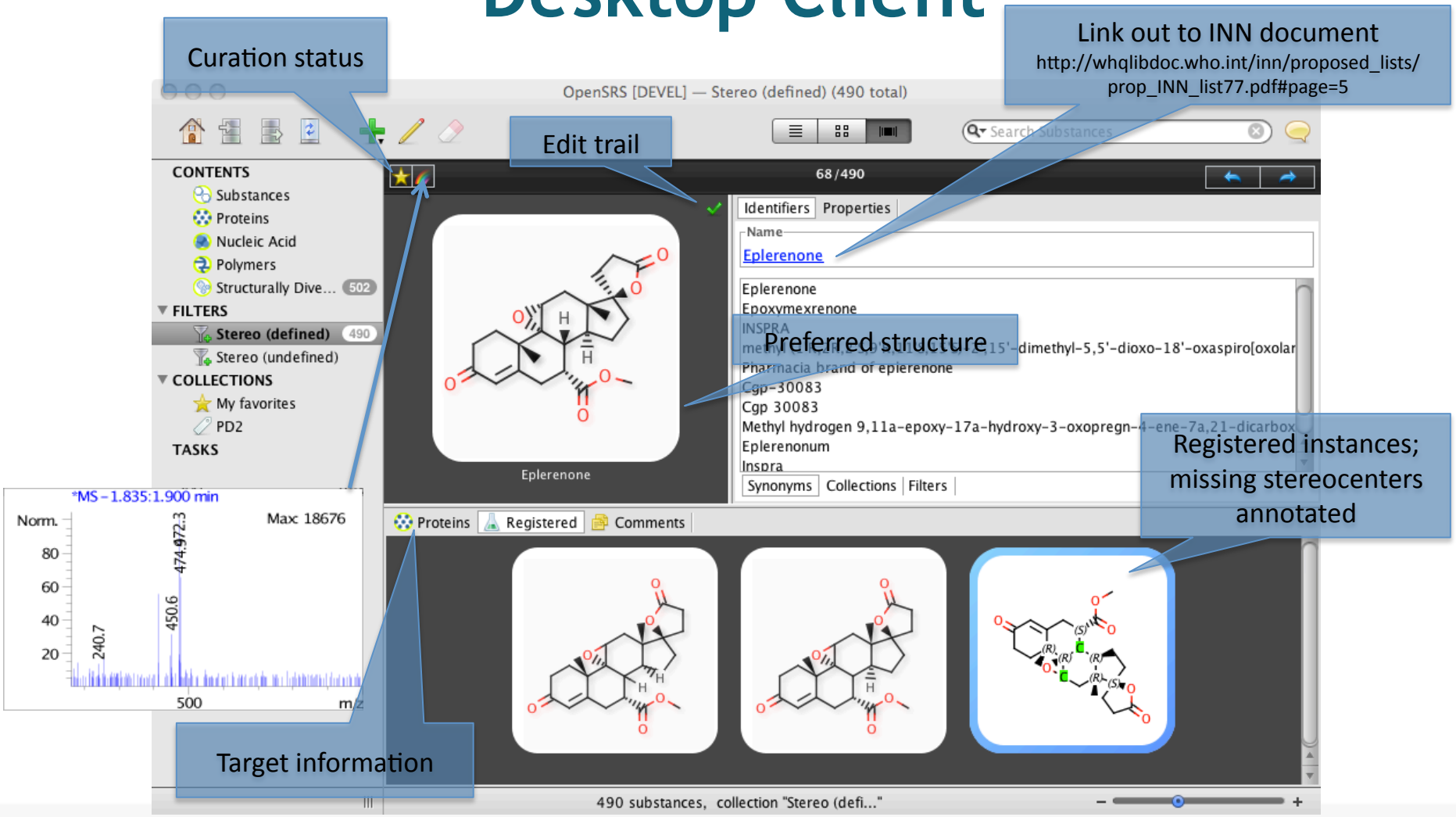
Web Client


OpenSRS
Search
Register
Approve
Administration
Enter New Substance
Entering
Similar (0)
Status
Substance Type
Chemical
Protein
Nucleic Acid
Polymer
Structurally Diverse
Mixture
Chemical Structure
New
Lasso
Single
Template
Nitrogen
Center
Ink
No JSDraw License Found
O
O
H2N
Stereochemistry
Type
ACHIRAL
Optical Activity
( )

Web Client


OpenSRS
Search
Register
Approve
Administration
Search Substance
By Name
By Structure
Advanced
Substructure
Superstructure
Similarity
Search
No JSDraw License Found
OH
Search Results

Desktop Client



Desktop Client